Bootstrap aggregated sparse FPCA for classification

Hyunsung Kim

Febrary 21, 2020

Department of Statistics
Chung-Ang University

Simulation studies 1

Simulation studies 1

- *Probability-enhanced effective dimension reduction for classifying sparse functional data* (Yao *et al.*)
- 3 simulation models

Model II: $f(x) = \exp(\langle \beta_1, X \rangle / 2) - 1,$

Model IV: $f(x) = \arctan(\pi \langle \beta_1, X \rangle) + \exp(\langle \beta_2, X \rangle / 3) - 1,$

Model New: $f(x) = \arctan(\pi \langle \beta_1, X \rangle / 4).$

- 700 curves are generated with 200 training and 500 test set.
- Bagged classifers are obtained from 100 bootstrap resamples.
- 100 Monte Carlo repetitions for each model

Simulation studies 1

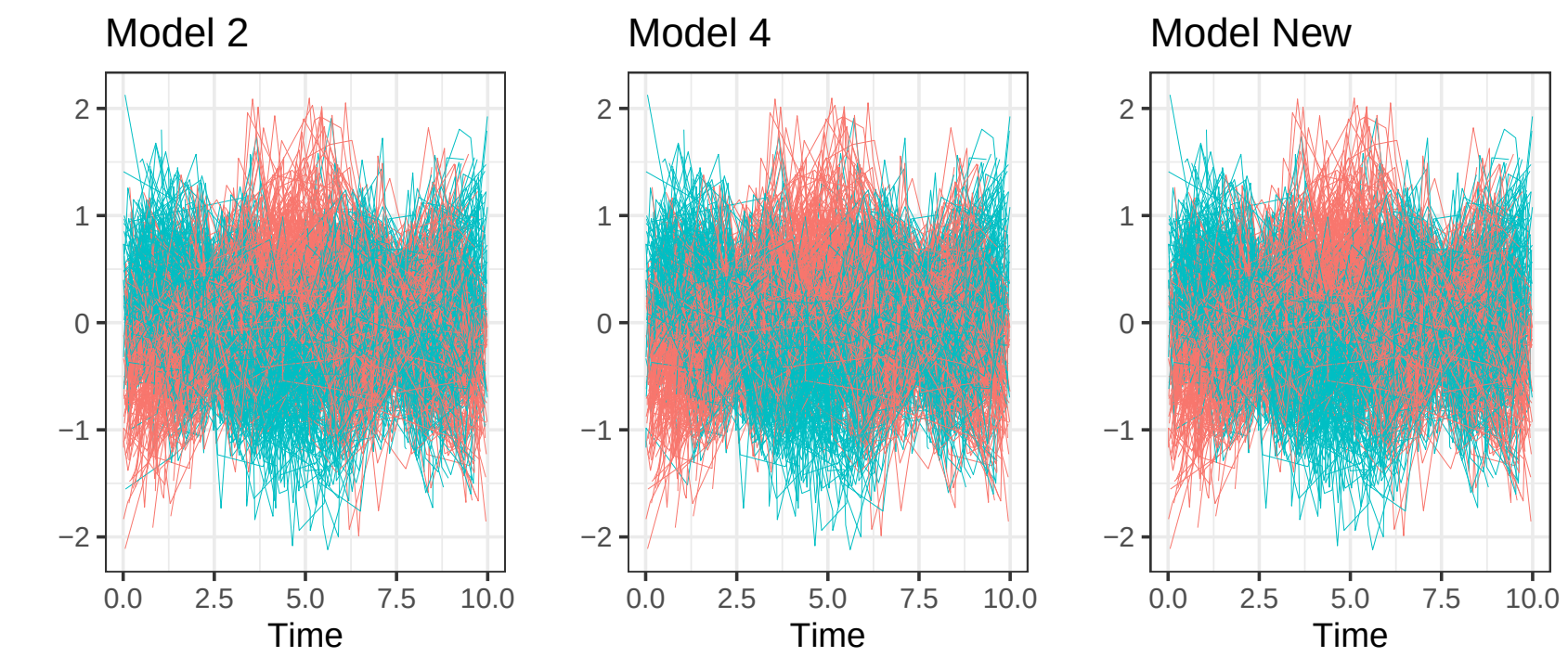

Figure 1: The simulated data obtained from 3 models

Results of simulation studies 1

Table 1: The average classification error with standard error in percentage from 100 Monte Carlo repetitions for 3 models

Model	Method	Logistic Regression	SVM (Linear)	SVM (Gaussian)	LDA	QDA	Naive Bayes
II	Single	16.7 (2.33)	16.8 (2.20)	17.5 (2.76)	16.6 (2.30)	17.8 (2.56)	18.4 (2.66)
	Majority vote	15.6 (1.95)	15.9 (1.87)	16.2 (2.28)	15.8 (1.96)	16.5 (2.14)	17.3 (2.42)
	OOB weight	16.0 (2.02)	16.2 (1.94)	16.6 (2.28)	16.1 (1.98)	16.9 (2.09)	17.7 (2.43)
IV	Single	12.8 (2.41)	12.8 (2.40)	13.3 (2.65)	12.8 (2.40)	13.8 (2.56)	14.8 (2.74)
	Majority vote	11.2 (1.84)	11.1 (1.89)	11.5 (1.98)	11.2 (1.85)	11.9 (2.03)	13.3 (2.36)
	OOB weight	11.6 (1.86)	11.5 (1.90)	12.0 (1.96)	11.6 (1.86)	12.3 (2.06)	13.6 (2.35)
New	Single	14.5 (2.17)	14.3 (2.18)	15.3 (2.69)	14.3 (2.17)	15.3 (2.36)	16.0 (2.22)
	Majority vote	13.1 (1.73)	13.1 (1.78)	13.6 (2.08)	13.1 (1.82)	13.8 (1.90)	14.9 (2.09)
	OOB weight	13.5 (1.81)	13.5 (1.78)	14.0 (2.03)	13.5 (1.84)	14.2 (1.92)	15.2 (2.12)

Simulation studies 2

Simulation studies 2

- Refer to *Functional Robust Support Vector Machines for Sparse and Irregular Longitudinal Data* (Wu & Liu)
- 3 simulation models

  Model A: Different mean and variance,

  Model B: Different mean,

  Model C: Different variance.

- 200 curves are generated with 100 training and 100 test set.
- Bagged classifers are obtained from 100 bootstrap resamples.
- 100 Monte Carlo repetitions for each model

Simulation studies 2

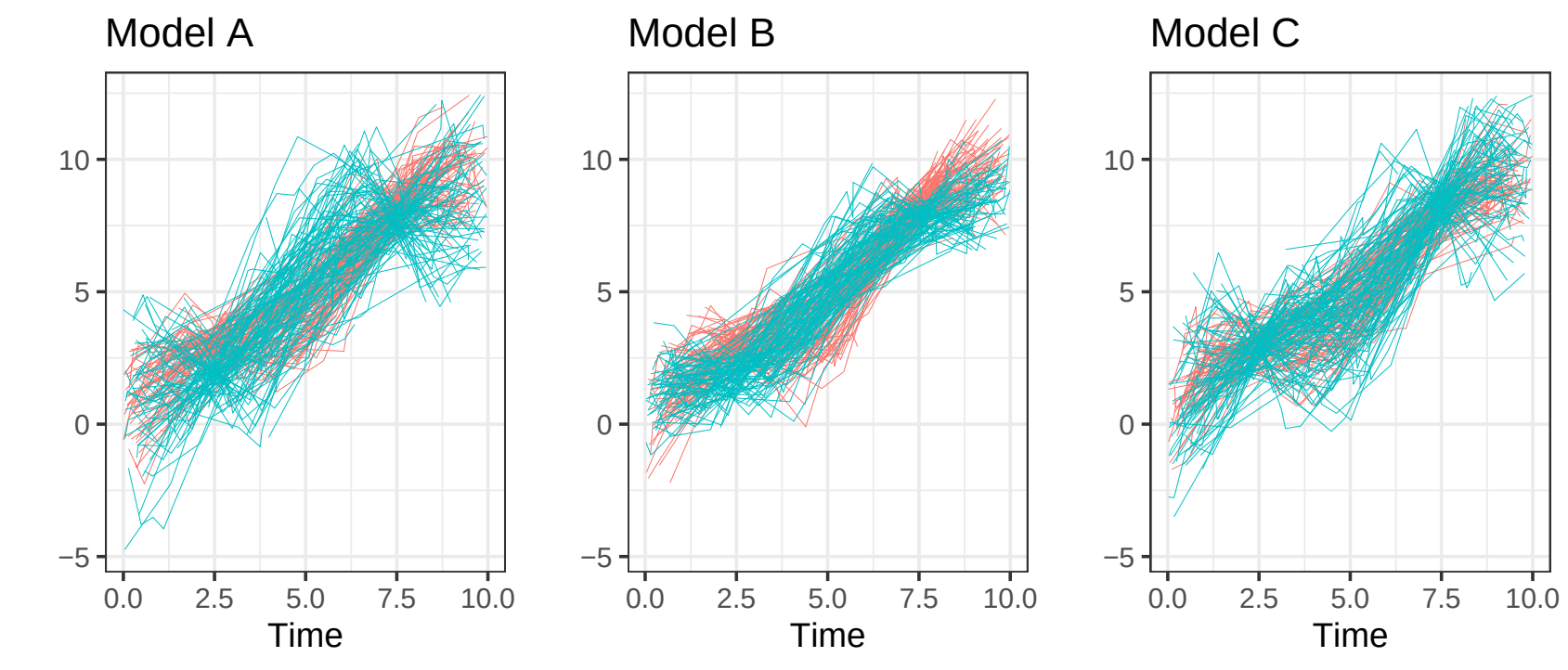

Figure 2: The simulated data obtained from 3 models

Results of simulation studies 2

Table 2: The average classification error with standard error in percentage from 100 Monte Carlo repetitions for 3 models

Model	Method	Logistic Regression	SVM (Linear)	SVM (Gaussian)	LDA	QDA	Naive Bayes
A	Single	17.6 (4.84)	17.5 (5.12)	15.3 (5.30)	17.1 (4.74)	15.1 (4.82)	16.5 (4.55)
	Majority vote	15.5 (4.23)	15.7 (4.41)	13.2 (4.35)	15.5 (4.08)	13.6 (4.17)	15.2 (4.09)
	OOB weight	16.1 (4.25)	16.3 (4.49)	13.9 (4.34)	16.3 (4.17)	14.2 (4.12)	15.7 (3.97)
B	Single	11.9 (3.43)	11.4 (3.49)	12.0 (4.25)	11.3 (3.55)	12.9 (3.62)	14.0 (4.39)
	Majority vote	10.7 (3.29)	10.4 (3.13)	11.0 (3.70)	10.4 (3.24)	11.6 (3.36)	12.4 (3.38)
	OOB weight	11.4 (3.27)	11.2 (3.00)	11.7 (3.60)	11.1 (3.30)	12.3 (3.36)	13.1 (3.59)
C	Single	50.5 (5.65)	49.5 (5.47)	32.8 (5.03)	50.6 (5.63)	31.2 (4.51)	30.5 (4.71)
	Majority vote	49.4 (5.58)	48.3 (6.15)	31.3 (5.09)	49.5 (5.68)	30.8 (4.06)	29.8 (4.29)
	OOB weight	48.9 (5.53)	47.8 (6.03)	31.5 (5.13)	48.8 (5.47)	31.0 (4.03)	30.1 (4.25)